

Sélection zygotique et tares récessives autosomiques dans une population finie

Gérard GÉRY

*U.E.R. Mathématiques, Math. 1^{er} cycle, Lyon I,
43, boulevard du 11-Novembre 1918,
69621 Villeurbanne*

Résumé

Cet article étudie le problème de la sélection zygotique dans une population finie relativement à une tare récessive autosomique.

Notre modèle stochastique met en évidence le rôle, sur le comportement de la fréquence aléatoire du gène porteur de la tare, des diverses formes de sélection rencontrées (sélection totale sur les homozygotes tarés, sélection partielle sur les homozygotes sains et les hétérozygotes).

La sélection totale des homozygotes tarés diminue la moyenne et la variance de la fréquence asymptotique du gène de tare a .

La sélection partielle peut, par un avantage sélectif des hétérozygotes Aa sur les homozygotes AA , rétablir la fréquence asymptotique moyenne du gène de tare a ; mais elle augmente sa dispersion.

A. — Introduction

Les techniques récentes de la médecine, plus particulièrement celles de l'amniocentèse, ouvrent de larges débouchés dans le domaine de la détection de la présence de tares chez les fœtus.

Des affections récessives comme la Maladie de NEIMANN-PICK, la Maladie de GAUCHER, la Mucopolysaccharidose sont reconnues par examen des cultures de fibroblastes prélevées dans le liquide amniotique. Cet examen révèle une déficience enzymatique spécifique ou la thésaurisation de substances biochimiques caractéristiques pour un désordre déterminé. L'avortement thérapeutique de fœtus atteints peut parfois se justifier et de toute façon exerce une action eugénique salutaire. Des analyses, assez grossières de l'effet de cette sélection ont été faites (STEELE et BREG, 1966; PFLANDER et KALIN, 1972). N'ayant pas accès au gamète, c'est un problème de sélection zygotique (CAVALLI-SFORZA, 1971).

Dans le domaine animal une étude analogue se pose pour l'éradication des gènes récessifs dans les populations bovines. Le progeny test, qui consiste, par

examen de la descendance de taureaux, à éliminer les hétérozygotes, est une forme de sélection zygotique (LEFORT et LAUVERGNE, 1974). Cette expérience concernant un groupe limité de descendants n'est pas d'une efficacité totale. L'élimination des hétérozygotes n'est que limitée, nous dirons que leur taux de survie n'est pas nul. D'autre part le phénomène de pénétrance incomplète conduit à envisager un taux de survie différent pour les 3 génotypes associés à un gène autosomal récessif.

Moins étudiée que la sélection gamétique, que l'on peut introduire sous forme linéaire dans un modèle stochastique (MALÉCOT, 1948), la sélection zygotique a surtout fait l'objet d'études en population infinie (JACQUARD, 1970).

Comme l'indiquent LEFORT et LAUVERGNE (1974), l'étude en population limitée introduit des fluctuations aléatoires non négligeables qu'il est indispensable d'analyser par le comportement en moyenne et surtout par la dispersion au voisinage de cette moyenne.

Notre but est l'étude en population finie de ce phénomène.

B. — Modèle stochastique

Considérons un modèle diploïde dans lequel tous les individus sont frères. Notre population diploïde est de taille finie N , à générations séparées et à sexes non séparés.

Notons : a le gène porteur de la tare et A son allèle;

$p(n)$, $q(n)$, $r(n)$ les fréquences aléatoires des diploïdes AA , Aa , aa à la n^e génération,

$t(n)$ la fréquence du gène a à la n^e génération dans la population.

La génération initiale D_0 est donnée par : p_0 , q_0 , r_0 .

Notons u , v , w les taux de viabilité des diploïdes-fœtus AA , Aa , aa .

Le passage de la n^e génération $D(n)$ à la $n + 1^e$ génération $D(n + 1)$ est réalisé de la manière suivante :

- tirage au hasard d'un diploïde X dans $D(n)$;
- tirage au hasard d'un diploïde Y dans un réservoir extérieur infini E de constitution fixée (AA et Aa en proportion $1 - \alpha$ et α);
- accouplement de X et Y et formation d'un réservoir infini R_{n1} de diploïdes-fœtus de constitution :

$$V_1 = (p_1(n + 1), q_1(n + 1), r_1(n + 1))$$

- sélection zygotique dans R_{n1} et formation d'un réservoir infini R_{n2} de diploïdes-fœtus de constitution :

$$V_2 = (p_2(n + 1), q_2(n + 1), r_2(n + 1))$$

- tirage au hasard de N diploïdes dans R_{n2} formant la $n + 1^e$ génération $D(n + 1)$.

Lorsque $D(n)$ est connu, les vecteurs aléatoires V_1 et V_2 suivent une loi de probabilité définie par :

$$P(V_1 = (1, 0, 0)) = P(V_2 = (1, 0, 0)) = (1 - \alpha)p(n),$$

$$P\left(V_1 = \left(\frac{1}{2}, \frac{1}{2}, 0\right)\right) = P\left(V_2 = \left(\frac{u}{u+v}, \frac{v}{u+v}, 0\right)\right) = \alpha p(n) + (1 - \alpha)q(n),$$

$$\begin{aligned} P\left(V_1 = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)\right) &= P\left(V_2 = \left(\frac{u}{u+2v+w}, \frac{2v}{u+2v+w}, \frac{w}{u+2v+w}\right)\right) = \alpha q(n), \\ P(V_1 = (0, 1, 0)) &= P(V_2 = (0, 1, 0)) = (1 - \alpha)r(n), \\ P\left(V_1 = \left(0, \frac{1}{2}, \frac{1}{2}\right)\right) &= P\left(V_2 = \left(0, \frac{v}{v+w}, \frac{w}{v+w}\right)\right) = \alpha r(n). \end{aligned}$$

C. — Équations de récurrence

L'étude des moments a priori (connaissance de la génération initiale D_0) est obtenue en utilisant le théorème des moyennes conditionnées. La moyenne a priori est égale à la moyenne de moyennes conditionnées (par la connaissance de la n^e génération D), nous avons :

$$E_{R_{n_2}}(\phi(n+1)) = \phi_2(n+1),$$

$$E_{D(n)}(E_{R_{n_2}}(\phi(n+1))) = \left(1 - \alpha \frac{v}{u+v}\right) \phi(n) + \frac{u}{u+v} \left(1 - \alpha \frac{v+w}{u+2v+w}\right) q(n),$$

$$E(\phi(n+1)) = \overline{\phi(n+1)} = \left(1 - \alpha \frac{v}{u+v}\right) \overline{\phi(n)} + \frac{u}{u+v} \left(1 - \alpha \frac{v+w}{u+2v+w}\right) \overline{q(n)},$$

de même :

$$\overline{q(n+1)} = \alpha \frac{v}{u+v} \overline{\phi(n)} + \frac{v}{u+v} \left(1 - \alpha \frac{w-u}{u+2v+w}\right) \overline{q(n)} + \left(1 - \alpha \frac{w}{v+w}\right) \overline{r(n)}$$

$$\overline{r(n+1)} = \alpha \frac{w}{u+2v+w} \overline{q(n)} + \alpha \frac{w}{v+w} \overline{r(n)}$$

posons :

$$X(n) = \begin{bmatrix} \phi(n) \\ q(n) \\ r(n) \end{bmatrix} \quad \overline{X(n)} = E(X(n)) \quad \overline{X} = \lim_{n \rightarrow +\infty} \overline{X(n)} \quad \overline{t} = \lim_{n \rightarrow +\infty} \overline{t(n)}$$

nous obtenons :

$$\overline{X(n+1)} = A \overline{X(n)}$$

$$A = \begin{bmatrix} 1 - \alpha \frac{v}{u+v} & ; & \frac{u}{u+v} \left(1 - \alpha \frac{v+w}{u+2v+w}\right) & ; & 0 \\ \alpha \frac{v}{u+v} & ; & \frac{v}{u+v} \left(1 - \alpha \frac{w-u}{u+2v+w}\right) & ; & 1 - \alpha \frac{w}{v+w} \\ 0 & ; & \alpha \frac{w}{u+2v+w} & ; & \alpha \frac{w}{v+w} \end{bmatrix}$$

La matrice A admet 3 valeurs propres réelles, ce sont les racines de l'équation :

$$(x - 1) \left(-x^2 + x \left(\frac{v}{u+v} + \alpha \left(\frac{w}{v+w} - \frac{2v(v+w)}{(u+v)(u+2v+w)} \right) \right) + \alpha(1 - \alpha) \frac{w(uw - v^2)}{(u+v)(v+w)(u+2v+w)} \right) = 0$$

avec : $s_1 = 1 \quad -1 < s_2, s_3 < 1.$

Si S est la matrice spectrale et P la matrice des vecteurs propres, nous obtenons :

$$\overline{X(n)} = PS^n P^{-1} X(0)$$

soit :

$$\overline{X} = K \begin{bmatrix} \frac{u}{u+v} \left(1 - \alpha \left(\frac{w}{v+w} + \frac{v+w}{u+2v+w} \right) + \alpha^2 \frac{w}{u+2v+w} \right) \\ \alpha \frac{u}{u+v} \left(1 - \alpha \frac{w}{v+w} \right) \\ \alpha^2 \frac{v}{u+v} \cdot \frac{w}{u+2v+w} \end{bmatrix}$$

$$\overline{t} = K \frac{\alpha}{2} \cdot \frac{v}{u+v} \left(1 + \alpha \frac{w(w-u)}{(v+w)(u+2v+w)} \right)$$

avec :

$$K^{-1} = \frac{u}{u+v} + \frac{\alpha}{u+v} \left(v - u \left(\frac{w}{v+w} + \frac{v+w}{u+2v+w} \right) \right) + \alpha^2 \frac{w(uw - v^2)}{(u+v)(v+w)(u+2v+w)}$$

La fonction génératrice des moments conjoints conditionnés par R_{n_2} de $p(n+1)$, $q(n+1)$, $r(n+1)$ est donnée par :

$$F(x, y, z) = (p_2 e^{\frac{x}{N}} + q_2 e^{\frac{y}{N}} + r_2 e^{\frac{z}{N}})^N.$$

nous avons :

$$E_{R_{n_2}}(r^2(n+1)) = \frac{\partial^2}{\partial z^2} F(0, 0, 0)$$

$$E(E_{D(n)}(E_{R_{n_2}}(r^2(n+1)))) = \overline{\alpha q(n)} \left[\frac{\partial^2 F(0, 0, 0)}{\partial z^2} \right]_{r_2 = \frac{w}{u+2v+w}} + \overline{\alpha r(n)} \left[\frac{\partial^2 F(0, 0, 0)}{\partial z^2} \right]_{r_2 = \frac{w}{v+w}}$$

$$\overline{r^2(n+1)} = \overline{\alpha q(n)} \frac{w}{(u+2v+w)^2} \left(w + \frac{u+2v}{N} \right) + \overline{\alpha r(n)} \frac{w}{(v+w)^2} \left(w + \frac{v}{N} \right)$$

de même :

$$\begin{aligned} \overline{q(n+1)r(n+1)} &= \alpha v w \left(1 - \frac{1}{N} \right) \left(\frac{\overline{r(n)}}{(v+w)^2} + 2 \frac{\overline{q(n)}}{(u+2v+w)^2} \right) \\ \overline{q^2(n+1)} &= (\alpha \overline{p(n)} + (1-\alpha) \overline{q(n)}) \left(v + \frac{u}{N} \right) \frac{v}{(u+v)^2} \\ &+ \alpha \overline{q(n)} \frac{2v}{(u+2v+w)^2} \left(2v + \frac{u+w}{N} \right) + (1-\alpha) \overline{r(n)} + \alpha \overline{r(n)} \frac{v}{(v+w)^2} \left(v + \frac{w}{N} \right) \end{aligned}$$

soit :

$$\text{var } t(n) = \frac{1}{4} (\overline{q^2(n)} - \overline{q(n)}^2 + \overline{r^2(n)} - \overline{r(n)}^2) + \frac{1}{2} (\overline{q(n)r(n)} - \overline{q(n)} \cdot \overline{r(n)})$$

D. — Discussion

I. — *Sélection ne portant que sur les homozygotes aa* ($u = v = 1$)

Notons $\bar{t}_0 = \bar{t}(w = 0)$, la valeur de \bar{t} lorsqu'il y a sélection totale sur les homozygotes tarés aa ($w = 0$).

$\bar{t}_1 = \bar{t}(w = 1)$, lorsqu'il n'y a pas de sélection ($w = 1$).

L'effet de cette sélection se traduit par une baisse de la moyenne asymptotique de $t(n)$.

$$\bar{t}_1 = \frac{\alpha}{2} \quad \bar{t}_0 = \frac{\alpha/2}{1 + 2\alpha/3}$$

Cette réduction, dont le taux est égal à $\frac{2\alpha/3}{1 + 2\alpha/3}$, peut ne pas être négligeable lorsque la fréquence des hétérozygotes Aa dans la population extérieure n'est pas faible (si $\alpha = 10$ p. 100 il vaut 6 p. 100).

L'évolution en moyenne de $t(n)$ dans les cas de sélection totale ($\overline{t_0(n)}$) et de non sélection ($\overline{t_1(n)}$) nous montre que :

$$\begin{aligned} \overline{t_0(n)} &= \frac{1}{2(1 + 2\alpha/3)} \left(\alpha - \frac{(1 - 2\alpha/3)^{n-1}}{2^n} (2(p_0 - 1) + q_0 + \frac{\alpha}{3}(p_0 + q_0 + 2)) \right. \\ &\quad \left. - \frac{2}{3} \alpha^2 (p_0 + q_0/3) \right) \end{aligned}$$

$$\overline{t_1(n)} = \frac{\alpha}{2} - \left(\frac{1}{2} \right)^{n+1} (2(p_0 - 1) + q_0 + \alpha(1 - q_0))$$

$\overline{t_0(n)}$ converge vers sa valeur limite comme $\left(\frac{1}{2} - \frac{\alpha}{3} \right)^n$ converge vers 0.

$\overline{t_1(n)}$ converge vers sa valeur limite comme $\left(\frac{1}{2} \right)^n$ vers 0.

La sélection par élimination des tarés *aa* augmente donc la vitesse de convergence de la fréquence du gène *a* (en moyenne).

Lorsque la population initiale D_0 n'est pas formée que d'homozygotes *AA* ($p_0 \neq 1$), la différence $\overline{t_1(n)} - \overline{t_0(n)}$ décroît et converge vers sa valeur limite $\frac{\alpha^2/3}{1 + 2\alpha/3} \cdot n \geq 1$.

Lorsque la population initiale D_0 est formée que d'homozygotes *AA* ($p_0 = 1$), alors :

$$\overline{t_0(n)} = \frac{\alpha}{2(1 + 2\alpha/3)} \left(1 - \left(\frac{1 - \alpha}{2} \right)^n \right),$$

$$\overline{t_1(n)} = \frac{\alpha}{2} \left(1 - \left(\frac{1}{2} \right)^n \right).$$

La différence $\overline{t_1(n)} - \overline{t_0(n)}$ croît et converge vers sa valeur limite $\frac{\alpha^2/3}{1 + 2\alpha/3}$.

Nous notons que :

$$\text{var}(t_0(\infty)) = \frac{\alpha/4}{(1 + 2\alpha/3)^2} \left(\frac{1}{2} \left(1 + \frac{1}{N} \right) - \frac{5\alpha}{9} \left(1 - \frac{2}{5N} \right) + \frac{2\alpha^2}{27} \left(1 - \frac{1}{N} \right) \right),$$

$$\text{var}(t_1(\infty)) = \frac{\alpha}{4} \left(\frac{1}{2} \left(1 + \frac{1}{N} \right) - \frac{13}{16} \alpha \left(1 + \frac{7}{13N} \right) + \frac{13}{32} \alpha^2 \left(1 + \frac{3}{13N} \right) - \frac{\alpha^4}{32} \right).$$

La sélection totale ($w = 0$) provoque une diminution de la variance asymptotique de $t(n)$.

(Si $\alpha = 10$ p. 100, $N = 100$, $\text{Var}(t_0(\infty)) = 0,0098$, $\text{Var}(t_1(\infty)) = 0,0106$, le taux de diminution de la variance asymptotique est de 7,5 p. 100.

2. — *Sélection totale* ($w = 0$) *sur les homozygotes aa, sélection partielle sur les homozygotes AA* ($0 < u < 1$) *et les hétérozygotes Aa* ($0 < v < 1$)

Posons : $x = \frac{u}{v}$,

nous obtenons : $\bar{t}_2 = \frac{\alpha/2}{x + \frac{2\alpha}{2+x}}$ avec $\bar{t}_2 = \bar{t}(w = 0; \frac{u}{v} = x)$

nous remarquons que \bar{t}_2 est une fonction décroissante de x .

Un avantage sélectif des hétérozygotes *Aa* par rapport aux homozygotes *AA* se traduira par une diminution de x et aura pour effet d'augmenter t_2 .

Un avantage sélectif relatif égal à $\frac{2}{(9 - 8\alpha)^{1/2} - 1} - 1$ permettra de maintenir l'état asymptotique de la population lorsqu'il n'y a aucune sélection ($\bar{t}_1 = \frac{\alpha}{2}$) (si $\alpha = 10$ p. 100 $\frac{v-u}{u} = 7$ p. 100), ce résultat confirme les résultats expérimentaux de ALLISON (1964) sur l'anémie falciforme.

Nous notons que :

$$\text{var } (t_2(\infty)) = \frac{\alpha}{4x(1+x)^2 \left(1 + \frac{2\alpha}{x(2+x)}\right)} \left((1+x) \left(1 + \frac{x}{N}\right) + \alpha \left(\frac{2x(1+x)}{(2+x)^2} \left(1 - \frac{1}{N}\right) - \frac{1}{x + \frac{2\alpha}{2+x}} \right) \right)$$

Var ($t(\infty)$) est une fonction décroissante de x , par suite un avantage sélectif des hétérozygotes se traduira par une dispersion plus grande.

Si α est petit, nous avons :

$$\text{Var } (t_2(\infty)) = \frac{1 + \frac{x}{N}}{4x(1+x)} + o(\alpha)$$

E. — Conclusion

La sélection par élimination des homozygotes tarés ($w = 0$) en génétique humaine (avortement thérapeutique) joue, en moyenne, un rôle sur la présence du gène porteur de la tare décroissant au cours de l'évolution des générations.

La diminution asymptotique de la moyenne de la fréquence du gène tare ($\bar{t}_0 = \frac{\alpha/2}{1 + 2\alpha/3}$ au lieu de $\bar{t}_1 = \frac{\alpha}{2}$) pouvant ne pas être négligeable (si $\alpha = 10$ p. 100, $\frac{\bar{t}_1 - \bar{t}_0}{\bar{t}_1} = 6$ p. 100).

Cette sélection accélère la convergence vers cet état asymptotique et diminue la dispersion d'autant plus que l'effectif N est grand (si $\alpha = 10$ p. 100, $N = 100$, le taux de diminution de la variance asymptotique est de 7,5 p. 100).

En génétique animale la sélection par élimination des tarés homozygotes ($w = 0$) est accompagnée d'une sélection partielle résultant du progeny test et de l'effet de la pénétrance incomplète ($0 < u < 1$ et $0 < v < 1$).

Cette sélection partielle se traduira par un avantage sélectif ($x = \frac{u}{v}$).

Un avantage sélectif des hétérozygotes (diminution de x) a pour effet d'augmenter en moyenne la présence du gène tare.

$\left(t = \frac{\alpha/2}{x + \frac{2\alpha}{2+x}} \right)$ et peut maintenir l'état asymptotique sans sélection ($\alpha = 10$ p. 100, $\frac{1}{x} = 1,07$).

Cet avantage sélectif joue un rôle sur la dispersion, il augmente la dispersion asymptotique de la variable aléatoire représentant le taux du gène porteur de la tare ($\text{Var } (t_2(\infty)) \neq \frac{1 + x/N}{4x(1+x)} \alpha$).

Le progeny test (élimination partielle des hétérozygotes, augmentation de x), réduit en moyenne la présence de la tare, mais diminue la dispersion asymptotique d'autant plus que N est grand.

Summary

Zygotic selection and autosomal-recessive affection in a finite population

This work studies the problem of zygotic selection in a finite population for an autosomal-recessive affection.

This problem of selection exists in therapeutic abortion and in the eradication of recessive genes in the bovine population (progeny-test). We use the following stochastic model:

A brother-diploid population, of size N , with separated generation, with non-separated sexes.

The $(n + 1)^{th}$ generation $D(n + 1)$ is obtained by the random drawing of a diploid X from the n^{th} generation $D(n)$, and a diploid Y from an infinite bowl of determined constitution (α proportion of Aa , $1 - \alpha$ proportion of AA).

A coupling of X and Y , after zygotic-selection at a fetal level provides by random drawing the $(n + 1)^{th}$ diploid generation $D(n + 1)$.

u, v, w are the frequencies of the n^{th} fetus-diploids generation AA, Aa, aa .

The formula for conditional expectations gives the a priori moments of the frequency of the gene to the n^{th} generation: $t(n)$.

The non-selection ($u = v = w = 1$) shows that $\overline{t_1(n)}$ converges to the asymptotic value $\overline{t_1} = \alpha/2$ as $(1/2)^n$ to 0.

The total selection of affected-homozygotes ($u = v = 1, w = 0$) shows that $\overline{t_0(n)}$ converges to its asymptotic value $\overline{t_0} = \frac{\alpha/2}{1 + 2\alpha/3}$ as $(1/2 - \alpha/3)^n$ to 0.

This selection increases the speed of convergence for $\overline{t(n)}$ and decreases its asymptotic value, it causes a diminution of the asymptotic variance to $t(n)$.

The selection of affected homozygotes ($w = 0$) and a partial selection of heterozygotes and non-affected homozygotes ($0 < u < 1, 0 < v < 1$) shows that $\overline{t(n)}$ is a decreasing function of $x = u/v$ ($\overline{t_2} = \overline{t}(w = 0, x) = \frac{\alpha/2}{x + \frac{2\alpha}{2+x}}$).

An selective advantage of the Aa heterozygotes in comparison with the AA homozygotes ($\frac{v-u}{u} = 7$ p. 100, if $\alpha = 10$ p. 100) allows the average maintaining of the asymptotic value of the frequency which corresponds to the non-selected situation, but increases the dispersion

$$\text{var } t_2(\infty) = \frac{1 + x/N}{4x(1 + x)}$$

Références bibliographiques

- ALLISON A. C., 1964. *Polymorphism and Natural Selection in Human Populations*. Cold Spring Harbor Symposia on Quantitative Biology, **XXIX**, 137-149.
- CAVALLI-SFORZA, 1971. *The Genetics of Human Populations*. Freeman, San Francisco.
- GÉRY G., 1976. Effet sélectif de l'élimination des tarés pour une tare liée au sexe dans une population finie. *Bio-Mathématique*, **55**, 137-151.
- JACQUARD A., 1970. *Structures Génétiques des Populations*. Masson, Paris.
- LÉFORT G., LAUVERGNE J. J., 1974. L'éradication des gènes récessifs à effets visibles dans des populations bovines où les taureaux sont mis à l'épreuve sur leur descendance *C.R. Acad. Sci. Paris*, **279** (D), 1381-1384.
- MALÉCOT G., 1966. *Probabilités et Hérité*. P.U.F., Paris.
- PFLANDER U., KALIN A., 1972. L'effet sélectif de l'avortement thérapeutique. *J. Génét. hum.*, **20**, 135-150.
- STEELE M. W. and BREG W. R., 1966. *Chromosome analysis of human amniotic-fluid*. *Lancet*, 383-385.